Compressed sensing for graphs

Seeun Umboh
1210 W. Dayton St., Madison, WI 53706-1685.

Abstract

We explore the possibility of using the recent Exact Matrix Completion(EMC) algorithm to reconstruct a graph using only a small number of its edges. Graphs model a wide variety of data such as social networks and computer networks, and we would like to reconstruct these networks using a small number of queries about the graph. We look at different types of graphs and analyze the effectiveness of the EMC algorithm in reconstructing them.

Introduction

We are interested in the following problem: given a graph and the ability to query whether an edge is in the graph or not, reconstruct the graph using as few queries as possible. For example, advertising companies may be interested in influential people. One way to find them is to look at individuals with high node degree in a social network, say Facebook. However, in general, obtaining information about edges can be time-consuming or access to the information is possibly restricted. So, we are interested in using significantly less than n^2 queries, and then applying a reconstruction algorithm to recover information about the unknown edges.

Recently, (Candes and Recht 2008) showed that we can recover a $n \times n$ matrix with low rank r using $O(n^{1.2} r \log n)$ random samples of the matrix entries using nuclear norm minimization[1]. The Singular Value Thresholding(SVT) algorithm was proposed by (Cai, Candes, and Shen 2008) and we use the implementation as given on *http://svt.caltech.edu*.

Exact matrix completion

We formally spell out the conditions as given in (Candes and Recht 2008) for exact matrix completion to be feasible.

Since we are interested in adjacency matrices, we will consider symmetric $n \times n$ matrices with 0 or 1 as entries. First of all, we need the matrix to be of low rank r. Ideally $r = O(1)$. But low rank is insufficient.

[1](Candes and Tao 2009) gives the better bound of $O(nr \log n)$ but the SVT algorithm is based on (Candes and Recht 2008).

Definition 1. *Let U be a subspace of $\mathbb{R}^n$ of dimension r and P_U be the orthogonal projection onto U. Then the coherence of U(with respect to the standard basis (e_i)) is defined to be*

$$\mu(U) = \frac{n}{r} \max_{1 \le i \le n} ||P_U e_i||^2.$$

Say M is a $n \times n$ matrix whose SVD is given by $M = USV^\top$ (Candes and Recht 2008) showed that if the matrix satisfies the following, then with high probability, a random sample of size $\Omega(n^{1.2} r \log n)$ of M's entries suffice to recover the matrix using nuclear norm minimization.

1. The coherences obey $\max(\mu(U), \mu(V)) \le \mu_0$ for some positive μ_0.
2. The maximum entry of M is bounded by $\mu_1 \sqrt{r}/n$ in absolute value for some positive μ_1.

Clearly, adjacency matrices satisfy the second condition. So, we only need to check for the first.

Method and results

We identified several graphs which have low rank and then classified them by whether their adjacency matrices are incoherent. Then, we ran the SVT algorithm, with various sample sizes and looked at the accuracy of the completion. For the experiments, we used Matlab with the above-mentioned SVT code. We also used the Matgraph package(*http://www.ams.jhu.edu/~ers/matgraph/*) to create graphs.

Many of the natural graph types turn out to be high rank. For example, trees, cycles and grids. Grids are interesting because (Kleinberg 2000) models social networks by overlaying random edges over a grid.

Stars, which are graphs with a center vertex that all other vertices are connected only to, have low rank but are highly coherent. Their matrices are of the form

$$\begin{bmatrix} 0 & 1 & 1 & 1 & 1 & \cdots \\ 1 & 0 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 0 & 0 & \cdots \\ \vdots & \ddots & & & & \end{bmatrix}$$

The vector $(1, 0, 0, 0, \cdots)$ is exactly in the row subspace and thus the star graph's matrix is coherent with the standard basis. As the theory predicts, the reconstruction is poor. All of its non-zero entries do correspond to edges in the graph. But it missed most of the edges.

An interesting graph type is the bipartite graph. Its matrix M_A is of the form

$$\begin{bmatrix} 0 & A \\ A^\top & 0 \end{bmatrix},$$

where A is allowed to be any arbitrary $0-1$ matrix. Its rank is $2\,\mathrm{rank}(A)$. Its coherence is also proportional to the coherence of A.

Bipartite graphs can be used to model preference lists and if we further allow multiple edges between any pair of vertices, we can model rating lists. For example, the left partition could be *users* and the right be *movies*. Then, the number of edges between a user and a movie indicates his rating of the movie. Typically, we simply perform the sampling on A instead of M_A. We expect that A is low-rank, because most people have similar tastes in movies, and so M_A should be low-rank too.

Surprisingly, complete bipartite graphs, which have rank 2 and very low coherence, cause the algorithm to diverge. Complete bipartite graphs are of the form $\begin{bmatrix} 0 & \mathbf{1} \\ \mathbf{1} & 0 \end{bmatrix}$, where $\mathbf{1}$ is the all-1 matrix. These have the lowest coherence among all bipartite graphs. However, reconstructing $\mathbf{1}$ poses no problem to the algorithm. All the entries in the reconstructed matrix are between 0.9 and 1.

Next, we tried running the algorithm on randomly generated bipartite graphs and randomly generated graphs. For the latter, we used the Erdos-Renyi model of random graphs where each edge is in the graph with equal probability and independently. However, on all our 50 trials, the rank of the graph is very high(close to n the number of vertices, and the dimension of the adjacency matrix). The algorithm either had a high relative error or diverged, where relative error is defined as

$$||X - M||_F / ||M||_F,$$

with $||\cdot||_F$ being the Frobenius norm and X is the reconstruction and M is the matrix we are trying to recover.

Conclusion

Our results seem to indicate that matrix completion as implemented by SVT is ill-suited for graph reconstruction. A lot of natural graph types have high rank, random graphs also have high rank with high probability(as suggested by the experiments). The algorithm performs badly even for bipartite graphs with low rank and low coherence. The likeliest reason is that there is some reason specific to the SVT implementation of matrix completion that causes its poor performance on the task of graph reconstruction.

References

Cai, J.; Candes, E.; and Shen, Z. 2008. A singular value thresholding algorithm for matrix completion. *Arxiv preprint arXiv:0810.3286*.

Candes, E., and Recht, B. 2008. Exact matrix completion via convex optimization. *preprint*.

Candes, E., and Tao, T. 2009. The Power of Convex Relaxation: Near-Optimal Matrix Completion. *Arxiv preprint arXiv:0903.1476*.

Kleinberg, J. 2000. The small-world phenomenon: an algorithm perspective. In *STOC '00: Proceedings of the thirty-second annual ACM symposium on Theory of computing*, 163–170. New York, NY, USA: ACM.